Time Series Analysis

Quantitative Investing I

Vrije Universiteit Amsterdam

Name:	Hamza Ghoussy, Andreas Koulopoulos, Milan Peter and Shubham Vijay Nanewar
Student Nr.:	2763379, 2897428, 2868506 and 2891203
Faculty:	School of Economics and Business
Supervisor:	Prof. Martijn de Vries

1 Data Setup

We will forecast the equity risk premium with 14 predictor variables based on Welch and Goyal (2008). The predictor dataset contains quarterly data for the 14 predictor variables along with the returns used for estimation. The predictive regression model for the equity premium is given by

$$r_{t+1} = \alpha_i + \beta_i x_{i,t} + \varepsilon_{t+1} \tag{1}$$

where r_{t+1} is the return of the stock market index in excess of the risk-free interest rate, $x_{i,t}$ is the predictive variable of interest, and ε_{t+1} denotes the disturbance term. Furthermore, i indexes the forecast variable of interest.

2 Part 1 - In-sample Regression

In this section, we will start with an in-sample estimation of our 14 predictive regressors according to Equation (1). According to the instructions, certain predictor variables must be negated before estimation, such as `ntis`, `tbl`, `lty`, and `lagINFL`. We ran the linear regression for each variable using OLS to find estimated coefficients($\beta_{i's}$).To account for potential heteroskedasticity and autocorrelation in error terms, we employ heteroskedasticity and autocorrelation consistent (HAC) standard errors with a maximum lag of four, corresponding to quarterly data. We have done the hypothesis testing one-tail test with hypothesis as below

$$H_0 : \beta_i = 0$$
$$H_A : \beta_i > 0$$

Rejection of the null hypothesis provides evidence that the predictor i has positive predictive content for future returns. The result of our OLS is given in the following snapshot showcasing the alpha, beta, t-values, the p-value of one side, and the adjusted R^2.

Table 1: In-Sample Predictive Regression Results

Predictor	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$	***t*-value**	***p*-value**	**Adj.** R^2
logDP	-0.088167	-0.030437	-1.942219	0.973578	0.016924
logDY	0.073612	0.017121	1.196164	0.116186	0.003409
logEP	-0.057684	-0.026527	-1.574210	0.941868	0.009126
logDE	0.002897	-0.019532	-0.791994	0.785573	0.001283
svar	0.033445	-2.049709	-2.080848	0.980944	0.082302
b/m	0.047364	-0.057355	-2.089105	0.981321	0.019556
ntis	0.019438	0.244881	0.828412	0.203977	0.001007
tbl	0.023855	0.252437	1.547288	0.061310	0.002906
lty	0.027364	0.238809	1.508643	0.066107	0.001538
ltr	0.015629	-0.007022	-0.047504	0.518932	-0.002600
tms	0.010700	0.291527	0.767270	0.221697	-0.001305
dfy	0.049000	-2.982694	-2.154155	0.984074	0.036104
dfr	0.013748	1.364118	4.555114	0.000004	0.091917
lagINFL	0.017204	0.224930	0.297026	0.383304	-0.001832

3 Part 2 - Out-of-sample Regression

In this section, we perform an out-of-sample estimation of the predictive regression model using an expanding window approach meaning that after each forecast step, we re-estimate our model using all the data up to that point and then generate the next forecast. The total sample of T observations is divided into three parts: an in-sample estimation period consisting of the first $m = 20$ years (80 quarters), a holdout validation period of $p = 10$ years (40 quarters), and an out-of-sample forecasting period composed of the remaining $q = T - (m + p)$ observations.

1. At time $t = m+p$, the first model is estimated using observations $1, \ldots, t$. The fitted coefficients $\hat{\alpha}_{i,t}$ and $\hat{\beta}_{i,t}$ are then used to forecast the excess return one period ahead:

$$\hat{r}_{i,t+1} = \hat{\alpha}_{i,t} + \hat{\beta}_{i,t} x_{i,t}.$$

2. The sample is then expanded by one additional observation. The regression is reestimated on the data $1, \ldots, t+1$, and the next one-step forecast is computed:

$$\hat{r}_{i,t+2} = \hat{\alpha}_{i,t+1} + \hat{\beta}_{i,t+1} x_{i,t+1}.$$

3. This recursive process continues until the final observation $T - 1$, producing a sequence of q forecasts $\{\hat{r}_t\}_{t=m+p+1}^{T}$.

At each forecast origin, a new OLS model is fitted using all data up to that point. The estimated coefficients are therefore time-varying and evolve as the information set expands. The benchmark forecast is defined as the historical mean of excess returns up to period t:

$$\bar{r}_{t+1} = \frac{1}{t} \sum_{j=1}^{t} r_j.$$

Forecast performance is assessed through the out-of-sample R^2_{OS} statistic of Campbell and Thompson (2008), which compares the mean squared prediction error (MSPE) of the predictive model with that of the historical mean benchmark:

$$R^2_{\text{OS}} = 1 - \frac{\sum_{k=p+1}^{p+q}(r_{m+k} - \hat{r}_{m+k})^2}{\sum_{k=p+1}^{p+q}(r_{m+k} - \bar{r}_{m+k})^2}.$$

A positive R^2_{OS} implies that the predictive regression model produces an MSPE lower than the benchmark and thus adds an incremental forecast value.

Result:

The expanding window out-of-sample analysis indicates limited robustness in the predictive regressions. Among the 14 predictors, only `lagINFL`, `ltr` and `dfy` deliver positive values R^2_{OS}, suggesting marginal improvements over the historical mean benchmark. These variables, reflecting inflation and credit conditions, retain modest predictive content for future excess returns. However, the gains are economically negligible, with all R^2_{OS} values below 0.5%. Most predictors produce negative values, indicating inferior performance relative to the benchmark. Overall, the results confirm the difficulty of obtaining robust equity premium forecasts, as any incremental predictive power is statistically weak and economically insignificant.

4 Part 3 - Kitchensink Regression

The kitchensink regression includes all predictor variables within a single model of the following form:

$$r_{t+1} = \hat{\alpha}_t + \hat{\boldsymbol{\beta}}_t' \mathbf{x}_t + \varepsilon_{t+1},$$

where $\mathbf{x}_t$ denotes the vector of all predictor variables, and $\hat{\alpha}_t$ and $\hat{\boldsymbol{\beta}}_t$ are the estimated coefficients based on the information available up to time t. This specification allows all predictors to simultaneously contribute to forecasting future excess returns.

Using the expanding-window approach with one-period-ahead forecasts (r_{t+1}), the kitchensink model produces an out-of-sample R^2_{OS} of -0.27. This negative value indicates that the combined model performs worse than the historical mean benchmark, suggesting that the inclusion of all predictors amplifies estimation noise and reduces the accuracy of the forecast. Poor performance likely reflects overfitting and strong multicollinearity among predictors, leading to unstable coefficient estimates as the sample expands.

When the same model is estimated using returns (r_t) instead of future returns, the out-of-sample R^2_{OS} increases to approximately 0.99. However, this reflects an in-sample fit rather than genuine predictive power, since the model effectively uses data already incorporated in the estimation.

In general, the results indicate that the kitchen sink regression does not improve the forecast performance. Including all predictors jointly increases model complexity without improving predictive accuracy, resulting in weaker out-of-sample performance compared to both the benchmark and single-variable models.

5 Part 4 - Forecast Combination Methods

To improve predictive accuracy, individual out-of-sample forecasts can be combined into a single aggregated forecast. The combined forecast for period $t+1$ is given by:

$$\hat{r}_{c,t+1} = \sum_{i=1}^{N} \omega_{i,t} \hat{r}_{i,t+1},$$

where $\hat{r}_{i,t+1}$ denotes the individual forecast from predictor i, and $\omega_{i,t}$ represents the ex-ante combination weight assigned to that forecast. The weights can be fixed or time-varying depending on the combination approach. Three combination methods are considered. First, the *mean combination forecast* assigns equal weights to all individual forecasts:

$$\omega_{i,t} = \frac{1}{N}, \quad i = 1, \ldots, N.$$

Second, the *median combination forecast* selects the median of individual forecasts $\{\hat{r}_{i,t+1}\}_{i=1}^{N}$, assigning unit weight to the median forecast and zero otherwise. Third, the *discounted mean squared prediction error (DMSPE)* method, determines weights based on past forecast accuracy:

$$\omega_{i,t} = \frac{\Phi_{i,t}^{-1}}{\sum_{j=1}^{N} \Phi_{j,t}^{-1}}, \quad \text{where} \quad \Phi_{i,t} = \sum_{s=m}^{t-1} \theta^{t-1-s} (r_{s+1} - \hat{r}_{i,s+1})^2,$$

and $\theta \in (0, 1]$ is a discount factor that places a greater weight on recent forecast performance. Lower values of $\Phi_{i,t}$ correspond to lower past forecast errors and therefore receive higher combination weights.
Result:
Table 2 reports the out-of-sample R^2_{OS} values for the forecast combination methods. The mean and median combinations slightly outperform the individual predictor forecasts, while the DMSPE-based method yields the strongest result.

Table 2: Out-of-Sample R^2_{OS} for Forecast Combination Methods

Combination Method	$R^2_{\mathbf{OS}}$
Mean Combination	0.0523
Median Combination	0.0064
DMSPE ($\theta = 1.0$)	0.0611
DMSPE ($\theta = 0.9$)	0.0791

Although all three combination methods improve the historical mean benchmark, their mechanisms and implications differ. The *mean combination* assigns equal importance to each prediction, effectively averaging idiosyncratic noise between predictors. It is simple and stable, but treats all predictors equally informative, ignoring differences in historical accuracy.

The *median combination* is more robust to outliers, as it relies on the central forecast rather than the mean. This makes it less sensitive to extreme individual predictions but also less responsive to genuine shifts in the underlying relationships when several predictors perform differently.

The *DMSPE combination* adapts dynamically by weighting forecasts according to past predictive performance. Predictors with lower mean squared prediction errors receive higher weights. When $\theta = 1$, all past errors are treated equally; when $\theta = 0.9$, the most recent performance receives greater weight. The latter captures the time variation in forecast accuracy more effectively, leading to the highest out-of-sample R^2_{OS} in our results.

In general, while the mean and median combinations provide baseline improvements through diversification, the DMSPE approach offers a more flexible and performance-driven weighting scheme that best balances adaptability and stability in forecasting.

6 Part 5 - Reflection on Findings

Comparison between in-sample and out-of-sample results reveals that variables showing strong in-sample significance do not necessarily translate into superior out-of-sample performance. Several predictors, such as `lagINFL`, `ltr`, and `dfy`, displayed statistically significant coefficients in-sample and remained among the few variables with positive R^2_{OS} values out-of-sample. However, their explanatory power decreased substantially once the models were evaluated in a true forecasting context. This indicates that apparent predictability in-sample is largely driven by sample-specific relationships that fail to persist over time.

From an economic perspective, forecast combination methods are expected to outperform individual models because they diversify model-specific errors and take into account different weighing schemes. The results support this expectation: all combination approaches produced positive out-of-sample R^2_{OS} values, with the DMSPE-based methods outperforming both the mean and median combinations. By dynamically assigning greater weight to more accurate predictors, the DMSPE approach adapts to changing market conditions, enhancing predictive stability and reducing overfitting.

When compared to the kitchensink regression, the forecast combinations deliver markedly better results. The kitchensink model, with an R^2_{OS} of approximately -0.27, underperformed due to multicollinearity and overfitting caused by estimating many correlated predictors simultaneously. In contrast, the combination forecasts achieved positive and economically meaningful R^2_{OS} values, confirming that aggregating information across individual models is more effective than jointly estimating all predictors.

Overall, the DMSPE combination with $\theta = 0.9$ appears the most appropriate method. It provides the best trade-off between adaptability and stability, yielding the highest out-of-sample performance while maintaining economic interpretability. This highlights that flexible, performance-based forecast weighting schemes can meaningfully improve the robustness of equity premium predictability compared with both individual-variable and full-information approaches.